InfiniBand In-Network Computing Technology and Roadmap

ISC’19 June 2019
InfiniBand In-Network Computing Technology and Roadmap

- Daniel Gruner, University of Toronto
- Ron Hawkins, San Diego Supercomputing Center
- Richard Gram, Mellanox
- Gilad Shainer, Mellanox
Mellanox Accelerates Leading HPC and AI Systems

World’s Top 10 Supercomputers

1. Summit CORAL System
World’s Fastest HPC / AI System
9.2K EDR InfiniBand Nodes

2. Sierra CORAL System
#2 USA Supercomputer
8.6K EDR InfiniBand Nodes

3. Wuxi Supercomputing Center
Fastest Supercomputer in China
41K InfiniBand Nodes

5. 8K HDR InfiniBand Nodes

6. Fastest HPC / AI System in Japan
1.1K EDR InfiniBand Nodes

8. 1.5K EDR InfiniBand Nodes

© 2019 Mellanox Technologies
HDR InfiniBand Wins Next Generation Supercomputers
(Examples)

23.5 Petaflops
8K HDR InfiniBand Nodes
Fat Tree Topology

3.1 Petaflops
1.8K HDR InfiniBand Nodes
Fat-Tree Topology

1.7 Petaflops
2K HDR InfiniBand Nodes
Dragonfly+ Topology

1.6 Petaflops
HDR InfiniBand
hybrid CPU-GPU-FPGA
Fat-Tree Topology

World’s First
HDR InfiniBand
Supercomputer

India’s National
Supercomputing
Program

© 2019 Mellanox Technologies
The Need for Intelligent and Faster Interconnect

Faster Data Speeds and In-Network Computing Enable Higher Performance and Scale

**CPU-Centric (Onload)**

- Must Wait for the Data
- Creates Performance Bottlenecks

**Data-Centric (Offload)**

- Analyze Data as it Moves!
- Higher Performance and Scale
Accelerating All Levels of HPC / AI Frameworks

Application
- Data Analysis
- Real Time
- Deep Learning

Communication
- Mellanox SHARP In-Network Computing
- MPI Tag Matching
- MPI Rendezvous
- Software Defined Virtual Devices

Network
- Network Transport Offload
- RDMA and GPU-Direct RDMA
- SHIELD (Self-Healing Network)
- Enhanced Adaptive Routing and Congestion Control

Connectivity
- Multi-Host Technology
- Socket-Direct Technology
- Enhanced Topologies
Highest Performance and Scalability for Exascale Platforms

96% Network Utilization

7X Higher Performance

Flat Latency

2X Higher Performance

5000X Higher Resiliency

Deep Learning

HDR 200G

NDR 400G

XDR 1000G
Scalable Hierarchical Aggregation and Reduction Protocol (SHARP)
Scalable Hierarchical Aggregation and Reduction Protocol (SHARP)

- Reliable Scalable General Purpose Primitive
  - In-network Tree based aggregation mechanism
  - Large number of groups
  - Multiple simultaneous outstanding operations

- Applicable to Multiple Use-cases
  - HPC Applications using MPI / SHMEM
  - Distributed Machine Learning applications

- Scalable High Performance Collective Offload
  - Barrier, Reduce, All-Reduce, Broadcast and more
  - Sum, Min, Max, Min-loc, max-loc, OR, XOR, AND
  - Integer and Floating-Point, 16/32/64 bits
SHARP AllReduce Performance Advantages (128 Nodes)

SHARP enables 75% Reduction in Latency
Providing Scalable Flat Latency
SHARP AllReduce Performance Advantages
1500 Nodes, 60K MPI Ranks, Dragonfly+ Topology

SHARP Enables Highest Performance
Scalable Hierarchical Aggregation and Reduction Protocol
SHARP AllReduce Performance Advantages
Oak Ridge National Laboratory – Coral Summit Supercomputer

SHARP Enables Highest Performance
SHARP Performance Advantage for AI

- SHARP provides 16% Performance Increase for deep learning, initial results
- TensorFlow with Horovod running ResNet50 benchmark, HDR InfiniBand (ConnectX-6, Quantum)

P100 NVIDIA GPUs, RH 7.5, Mellanox OFED 4.4, HPC-X v2.3, TensorFlow v1.11, Horovod 0.15.0
Adaptive Routing
InfiniBand Proven Adaptive Routing Performance

- Oak Ridge National Laboratory – Coral Summit supercomputer
- Bisection bandwidth benchmark, based on mpiGraph
  - Explores the bandwidth between possible MPI process pairs
- AR results demonstrate an average performance of 96% of the maximum bandwidth measured

mpiGraph explores the bandwidth between possible MPI process pairs. In the histograms, the single cluster with AR indicates that all pairs achieve nearly maximum bandwidth while single-path static routing has nine clusters as congestion limits bandwidth, negatively impacting overall application performance.

GPUDirect
10X Higher Performance with GPUDirect™ RDMA

- Accelerates HPC and Deep Learning performance
- Lowest communication latency for GPUs

![Graph showing latency and throughput improvements with GPUDirect™ RDMA](image-url)
Network Topologies
Supporting Variety of Topologies

- Fat Tree
- Hypercube
- Torus
- Dragonfly
Traditional Dragonfly vs Dragonfly+
Dragonfly+ Topology

- Several “groups”, connected using all to all links
- The topology inside each group can be any topology
- Reduce total cost of network (fewer long cables)
- Utilizes Adaptive Routing for efficient operations
- Simplifies future system expansion

Full-Graph connecting every group to all other groups

1200-Nodes Dragonfly+ Systems Example
Dragonfly+ Topology

- Several “groups”, connected using all to all links
- The topology inside each group can be any topology
- Reduce total cost of network (fewer long cables)
- Utilizes Adaptive Routing for efficient operations
- Simplifies future system expansion

Full-Graph connecting every group to all other groups

1200-Nodes Dragonfly+ Systems Example
Future Expansion of Dragonfly+ Based System

- Topology expansion of a Fat Tree, or a regular/Aries like Dragonfly requires one of the following:
  - Reduction of early phase bisection bandwidth due to reservation of ports on the network switches
  - Re-cabling the long cables
- Dragonfly+ is the only topology that allows system expansion at zero cost:
  - While maintaining bisection bandwidth
  - No port reservation
  - No re-cabling
Re-cable the central racks, a change local to the RACK
DRAGONFLY+ AND CANADA’S FASTEST SUPERCOMPUTER NIAGARA
SCINET

• SciNet is a Consortium for High Performance Computing consisting of researchers at the University of Toronto and the associated Hospitals.

• Provides compute resources and support for Compute/Calcul Canada, Compute Ontario and SOSCIP.

• Is funded by the Canada Foundation for Innovation, NSERC, the Government of Ontario, Fed Dev Ontario, and the University of Toronto.
NIAGARA

DESIGN CRITERIA

• Designed for Large Parallel (MPI) jobs supporting academic research across Canada

• Competitive RFP Process
  o Application Benchmarks
  o Energy Efficiency (Flops/Watt)
  o Network Design
  o Storage Design
  o Deployment Plan

• 11 bidders, 5 shortlisted

• Lenovo bid chosen
NIAGARA

SYSTEM SPECIFICATIONS

- 1500 Lenovo SD530 compute nodes (2x20 core Intel Gold 6148 with 192GB Ram)
- 21 compute racks w/RDHx, 9 management racks
- 60,000 cores
- Mellanox ConnectX-5 EDR InfiniBand (Dragonfly+ Topology)
- 10PB spinning disk (GPFS)
- 256TB Burst Buffer (Excelero NVMe)
- Rpeak/Rmax of 4.61/3.07 (53rd on June 2018 Top500)
- 685 kW
NIAGARA

STORAGE
• 3x Lenovo DSS G260 (504x10TB)
• 10 PB Spinning Disk
• 70-90 GB/s R/W
• Spectrum Scale (GPFS)

BURST BUFFER
• 256 TB in Raid 1
• 10 Lenovo SR650 nodes (8x6.4TB NVMe)
• Excelero NVMe Fabric
• 160 GB/s R/W
• very high IOPs performance
• Spectrum Scale (GPFS)
NIAGARA

DRAGONFLY+

• Dragonfly topology first proposed around 2008, used in proprietary networks (Cray Aries & SlingShot)

• Dragonfly+ is a variant where groups are connected in a CLOS-like topology

• InfiniBand supported (ConnectX-5, Switch IB2), supported in OpenSM

• Performance
  o High bandwidth (determined by links between groups)
  o Latency (only two switch levels)

• Flexible & Cost-effective
  o Initial design: can be easily tailored to meet requirements (size of cluster, bandwidth)
  o Resizing: minimal invasiveness, ability to add groups, or fill out partial groups
  o Implemented using 1U switches and high number of copper cables

• Adaptive Routing

• Congestion Control
NIAGARA

DRAGONFLY+

• 1500 compute nodes connected in 4 groups (1728 connections without changes to core)

• Group (Level 1)
  o 432 nodes connected 1:1
  o (24 x 36 port switches) 18 copper to nodes, 18 optical to core

• Core (Level 2)
  o 4 groups of 18 ”core” 36 port 1U switches
  o 24 optical to group, 12 copper inter-core

• Total
  o 161 x SB7890 Switch IB-2 EDR
  o 1574 x ConnectX-5 EDR HCA’s
NIAGARA

DRAGONFLY+
NIAGARA

DRAGONFLY+
NIAGARA
NIAGARA

SOFTWARE

• Fabric Manager: UFM 5.11 with Adaptive Routing Enabled
• Centos 7.4 w/ Mellanox OFED 4.2 (soon to be OFED 4.4)
• HPC-X 2.1.0
• OpenMPI 3.1.0 (with SHARP, UCX, HCOLL) default MPI
Simulating Stars - Prof. Falk Herwig (U. of Victoria)

- Investigating how stellar convection, nuclear reactions, and atomic element formation processes work in the final stages of the lives of stars.
- 1536^3-grid case possible now in Canada using 1088 nodes (43.5k cores) in 60 hrs

Global Ocean Circulation - Prof. B. Arbic (U Mich), Dr. D. Menemenlis (NASA), Prof. W.R. Peltier (U of T)

- Attempting to understand how small-scale processes affect the global circulation of our oceans.
- 1200 node (48k core) simulations - 664 TB (3 months simulated)
NIAGARA
SDSC at a Glance

• Established 1985 as one of original NSF-funded supercomputer centers

• Transitioned from General Atomics to UC San Diego in 1997

• ~225 staff

• Large backlog of NSF awards (60+)

• >$1B cumulative revenues

• Currently operating 3 major HPC systems – Comet, TSCC, Gordon-Popeye

• Multiple proposals under review for new capabilities

Top photo: SDSC archives, Bottom Photo: B. Tolo
Comet -- Serving the “Long Tail” of HPC Users

Photo: J. Lou

This work supported by the National Science Foundation. award ACI-1341698.
Data extracted from NSF’s XDMoD data base informed a design that reflects the way researchers actually use HPC systems

- 99% of jobs run on NSF’s HPC resources in 2012 used <2,048 cores
- And consumed >50% of the total core-hours across all NSF resources

A system designed to serve the 99% is significantly different than one for the 1%.
Comet Built to Serve the 99%

CHALLENGES OUR PROPOSAL ADDRESSES
- Attract new users and communities
- Support diverse applications with complex workflows
- Ensure responsiveness for thousands of users
- Transfer, store, analyze, and share massive data sets
- Integrate with XSEDE

COMET COMPUTE SYSTEM
- Cluster architecture
  - Fast standard nodes
  - Large-memory nodes
  - GPU-accelerated nodes
  - FDR InfiniBand
- Storage architecture
  - Performance Storage
  - Durable Storage
- Software
  - Science Gateways
  - Rich base of installed apps
  - Virtualization

USER & SYSTEM SUPPORT
- New user orientation
- XSEDE collaborations
- FutureGrid

ALLOCATIONS & SCHEDULING
- Optimized for throughput
- Per-project allocation caps
- Per-job core limits

XSEDE Service Providers
- Internet2, ESnet @100G
  - (Universities, Labs)
- Open Science Grid

Science Gateways

Cluster

Cluster

Performance Storage

Durable Storage

UCSD Campus Bridging
(e.g., LHC Tier 2 Data Site)

- Island architecture
- Mix of node types
- Virtualized HPC clusters

SDSC
SAN DIEGO SUPERCOMPUTER CENTER

UC San Diego
Comet Network Architecture
InfiniBand compute, Ethernet Storage

Recently Upgraded to FDR IB

- **Node-Local Storage**: 320 GB, 72 HSWL, 4 Large-Memory
- **Core InfiniBand**: 2 x 108-port
- **IB-Ethernet Bridges**: 4 x 18-port each
- **Arista 40GbE**: (2x)
- **Research and Education Network Access Data Movers**
  - Juniper 100 Gbps
  - Arista 40GbE (2x)
  - Data Mover Nodes

**Additional Support Components**
(not shown for clarity)
- Ethernet Mgt Network (10 GbE)

**Performance Storage**
- 7.7 PB, 200 GB/s
- 32 storage servers

**Durable Storage**
- 6 PB, 100 GB/s
- 64 storage servers

**FDR 36p**
- 4 switches
- 18 racks

**FDR 72**
- 18 switches

**40GbE**
- 64
- 128
- 10GbE

7x 36-port FDR in each rack wired as full fat-tree. 4:1 over subscription between racks.
Comet: System Characteristics

• Total peak flops ~2.1 PF
  • Dell primary integrator
    • Intel Haswell processors w/ AVX2
    • Mellanox FDR InfiniBand
• 1,944 standard compute nodes (46,656 cores)
  • Dual CPUs, each 12-core, 2.5 GHz
  • 128 GB DDR4 2133 MHz DRAM
  • 2*160GB GB SSDs (local disk)
• 72 GPU nodes
  • 36 nodes same as standard nodes plus
    Two NVIDIA K80 cards, each with dual
    Kepler3 GPUs
  • 36 nodes with 2 14-core Intel Broadwell
    CPUs plus 4 NVIDIA P100 GPUs
• 4 large-memory nodes
  • 1.5 TB DDR4 1866 MHz DRAM
  • Four Haswell processors/node
  • 64 cores/node

• Hybrid fat-tree topology
  • FDR (56 Gbps) InfiniBand
  • Rack-level (72 nodes, 1,728 cores) full
    bisection bandwidth
  • 4:1 oversubscription cross-rack
• Performance Storage (Aeon)
  • 7.6 PB, 200 GB/s; Lustre
  • Scratch & Persistent Storage segments
• Durable Storage (Aeon)
  • 6 PB, 100 GB/s; Lustre
  • Automatic backups of critical data
• Home directory storage
• Gateway hosting nodes
• 100 Gbps external connectivity to
  Internet2 & ESNet
Comet Helps Confirm Gravitational Wave Discovery

COMET delivered 700,000 core-hours to LIGO

Signals in synchrony
When shifted by 0.007 seconds, the signal from LIGO’s observatory in Washington (red) neatly matches the signal from the one in Louisiana (blue).

- LIGO Hanford data (shifted)
- LIGO Livingston data

Strain (10^-20)
Time (seconds)
Atmospheric river forecasts

The West-WRF project (Plymouth State, UCSD) is using Comet to provide near-real time atmospheric river forecasts. Queue delays reduced to less than 100 seconds by setting a standing reservation.

Images provided by John Helly, Andrew Martin and Marty Ralph
AI + HPC in Science

Machine learning of accurate many-body potentials

Predictive simulations require accurate many-body potentials

Continuous analytical representation of highly complex quantum mechanical intermolecular interactions:
Polynomial expansions or neural networks

Machine Learning

Accurate but expensive: QM requires immense compute resources
e.g. CCSD(T)/CBS

Quantum mechanical reference interaction energy data points


Accurate and affordable: Much faster than QM
Enables long time-scale dynamics simulations with *ab initio* accuracy

*Slide Courtesy of Andreas Goetz, SDSC*
InfiniBand Developments
RDMA-Hadoop and RDMA-Spark

Collaboration with Dr. D K Panda’s Network-Based Computing Lab (Ohio State University)

• HDFS, MapReduce, and RPC over native InfiniBand and RDMA over Converged Ethernet (RoCE).

• Based on Apache distributions of Hadoop and Spark.

• Version RDMA-Apache-Hadoop-2.x 1.3.5 (based on Apache Hadoop 2.8.0) available on Comet

• Version RDMA-Spark 0.9.5 (based on Apache Spark 2.1.0) is available on Comet.

• More details on the RDMA-Hadoop and RDMA-Spark projects at:
  • http://hibd.cse.ohio-state.edu/
RDMA-Hadoop and Spark Design

- Exploit performance on modern clusters with RDMA-enabled interconnects for Big Data applications.
- Hybrid design with in-memory and heterogeneous storage (HDD, SSDs, Lustre).
- Keep compliance with standard distributions from Apache.
InfiniBand FDR, SSD, 32/64 Worker Nodes, 768/1536 Cores, (768/1536M 768/1536R)

- RDMA-based design for Spark 1.5.1
- RDMA vs. IPoIB with 768/1536 concurrent tasks, single SSD per node.
  - 32 nodes/768 cores: Total time reduced by 37% over IPoIB (56Gbps)
  - 64 nodes/1536 cores: Total time reduced by 43% over IPoIB (56Gbps)

Performance Evaluation on SDSC Comet – HiBench PageRank (Results from Dr. Panda’s group)
MVAPICH2-GDR on Comet

- Dr. DK Panda’s team (Ohio State), have been enhancing MVAPICH2 to utilize GPUDirect capabilities and achieve low MPI latencies and high bandwidth.

- MVAPICH2-GDR is available on the SDSC Comet GPU nodes featuring NVIDIA K80s and P100s.

- Applications that use MVAPICH2-GDR include HOOMD-Blue, OSU-Caffe, and TensorFlow (Horovod).
Lustre Ethernet vs IB

- 1 node with 24 ranks
- Data on /oasis/scratch/comet

- I/O performance with IB consistently better than with Ethernet
- Stable interactive I/O like file editing using 'vi'…
Triton Shared Computing Cluster
High Performance Computing for UC Researchers

- Mid-scale campus research cluster
- Launched 2013
- Hybrid business model: “condo” (buy in) and “hotel” (pay-as-you-go) options
- Mixed architecture:
  - ~375 nodes (>7,000 cores, not incl. GPU cores) + 50 GPU nodes (~300 GPUs with mix of NVIDIA GPUs)
  - 850TB parallel file system, home file system, add'l storage services
  - Mixed QDR/FDR/EDR + 10GbE
  - High bandwidth external networks
- 33 participating labs/groups

Photo: B. Tolo
Overall Architecture
TSCC BioBurst Detail
(NSF OAC Award #1659104)

**DDN IME System**

- **Home File System**: 40 TB
- **10GbE Switch**: To Lustre PFS via Arista 7508
- **Compute Nodes**:
  - 2 x 14-core Xeon (Broadwell)
  - 128GB DRAM
  - 2 x 480GB SATA SSD
  - 1 x EDR IB
  - 1 x 10GbE
- **EDR IB Switch**: Mellanox SB-7700 36-port managed
- **I/O Accelerator**
- **Core Ethernet Switches**: Arista 7508
- **Lustre Parallel File System**: 650TB

**New Capability**
- 100 Gb/s InfiniBand
- 40 Gb/s Ethernet
- 10 Gb/s Ethernet

**Nodes**:
- **DRAGEN Node**:
  - FPGA Accelerator
  - Custom Firmware
- **Additional Nodes**:
  - 2 x DDN IME-240 Appliance
  - 44TB x 1.2TB NVMe SSDs
  - 4 x EDR 100 Gbps IB
  - 2 x 40GbE

**DDN IME System**

- **Core Ethernet Switch**
- **New Capability**: 100 Gb/s InfiniBand
- **40 Gb/s Ethernet**
- **10 Gb/s Ethernet**
Computational/Big Data Resources for Cancer Research

Triton Shared Computing Cluster

NSF-funded BioBurst Cluster

Pursuing the Fight against Cancer
Professor Ludmil Alexandrov's research has been focused on understanding mutational processes in human cancer through the use of mutational signatures. In 2013, he developed the first comprehensive map of the signatures of the mutational processes that cause somatic mutations in human cancer. This work was published in several well-regarded scientific journals and highlighted by the American Society of Clinical Oncology as a milestone in the fight against cancer. More recently, Alexandrov mapped the signatures of clock-like mutational processes operative in normal somatic cells, demonstrated that mutational signatures have the potential to be used for targeted cancer therapy, and identified the mutational signatures associated with tobacco smoking.

Ludmil Alexandrov is an assistant professor with UC San Diego's Cellular & Molecular Medicine Department, and an Oppenheimer Fellow in the Theoretical Biology & Biophysics Group and the Center for Nonlinear Studies at Los Alamos National Laboratory.

Photo Credits: SDSC, UCSD
HDR InfiniBand
ConnectX-6 HDR InfiniBand Adapter

**Leading Connectivity**
- 200Gb/s InfiniBand and Ethernet
  - HDR, HDR100, EDR (100Gb/s) and lower speeds
  - 200GbE, 100GbE and lower speeds
- Single and dual ports

**Leading Performance**
- 200Gb/s throughput, 0.6usec latency, 215 million message per second
- PCIe Gen3 / Gen4, 32 lanes
- Integrated PCIe switch
- Multi-Host

**Leading Features**
- In-network computing and memory for HPC collective offloads
- Security – Block-level encryption to storage, key management, FIPS
- Storage – NVMe and NVMe-over-Fabrics accelerations
HDR InfiniBand Switch: QM8700, 1U Series

40 QSFP56 ports (50G PAM4 per lane)
- 40 ports of HDR, 200G
- 80 ports of HDR100, 100G

Superior performance
- 390M packets per sec (64B)
- 16Tb/s aggregate bandwidth

Superior resiliency
- 22” depth
- 6 fans (5+1), hot swappable
- 2 power supplies (1+1), hot swappable
HDR InfiniBand Switch: CS8500, Modular Series

800 QSFP56 ports
- 800 ports of HDR, 200G
- 1600 ports of HDR100, 100G

Superior performance
- 320Tb/s aggregate bandwidth
- LCD Tablet IO panel

Water-cooled solution
- Liquid – Liquid 4U CDU
- Liquid – Air 42U (350mm wide) stand alone HEX
- 0C – 35C (air) or 40C (water) operating air range
Real Time Network Visibility

Built-in Hardware Sensors for Rich Traffic Telemetry and Data Collection

Advanced monitoring for troubleshooting

- 8 mirror agents triggered by congestion, buffer usage and latency
- Measure queue depth using histograms (64ns granularity)

Network status/health in real time

- Buffer snapshots
- Congestion notifications and buffers status
InfiniBand Roadmap
InfiniBand Roadmap (IBTA)
Thank You